

A New Fast and Robust Method Based on Head Detection for People-Flow Counting System

Yaowu Hu, Ping Zhou, Hao Zhou

Institute of Vision Detection, Zhejiang Sci-Tech University, Hangzhou 310018, P.R. China

zp@zstu.edu.cn

Abstract-This paper presents an implementation method for the people counting system which detects and tracks moving people using a fixed single camera. The main contribution of this paper is the novel head detection method based on body's geometry. A novel body descriptor is proposed for finding people's head which is defined as Body Feature Rectangle (BFR). First, a vertical projection method is used to get the line which divides touching persons into individuals. Second, a special inscribed rectangle is found to locate the neck position which describes the torso area. Third, locations of people's heads can be got according to its neck-positions. Last, a robust counting method named MEA is proposed to get the real counts of walking people flows. The proposed method can divide the multiple-people image into individuals whatever people merge with each other or not. Moreover, the passing people can be counted accurately under the influence of wearing hats. Experimental results show that our proposed method can nearly reach to an accuracy of 100% if the number of a people-merging pattern is less than six.

Keywords-people counting; head detection; BFR; people-flow tracking

I. INTRODUCTION

People-flow counting system plays an important part in our security applications, such as tourist' flow estimation, traffic management, and supermarket management and so on. Generally speaking, People used sensors like light beams and rotary bar to count people-flow widely in the recent years. Although some methods based on image processing appeared, all of them had suffered several practical problems:

If many people are walking side by side, it is difficult to segment and count the accurate passing people.

When people are wearing various hats, they are difficult to be detected and counted.

The whole processing is time consuming.

To solve these problems, we develop a people-flow counting system using a new method for fast head detection based on BFR.

This paper is organized as follows. In section 2, the related work in people-flow counting is reviewed. In section 3, our architecture of people counting system is described in details. In section 4, some experimental results are shown. Conclusions and future work are drawn in section 5.

II. LITERATURE REVIEW

Existing approaches for people-flow counting in the surveillance area are generally classified into three categories:

- 1). Multiple-human image detection and segmentation method.
- 2). Model based training and classifying method.
- 3). Area estimation method.

The first approach is to divide a crowd of people image into individuals. Finding a robust algorithm of image segmentation is the key to success for the people counting system. Distance transformation algorithm [1] calculates the distance between foreground pixel and nearest horizontal neighbor background pixel to segment the multiple-human image. It can only acquire satisfactory results when two people have a little overlap with each other. Chen [4] used a novel method based on area and color analyses to overcome the overlap problems of touching persons. He labeled each people with a special color vector and tracked each people pattern with an analysis of its HSI histogram. This method is effective for lots of people expect many of them wearing in the same color. Liu and Tu [5] proposed a new model. In order to divide multiple-people image into several single people images, they found image-features using a likelihood function and estimated the position of people by the EM algorithm. This method is time consuming.

The second one is a model-based approach to find people's features by training data of images. For example, Hua and Lei [2] described a method which detects people by a special model of head-shoulder for normal people. The head-shoulder model is obtained through training and classifying thousands of samples by the linear SVM. Liu [12] detected people's head by the Hough transform algorithm. It is complex and imprecise.

The third approach usually counts the number of people by some estimation algorithms. Ye and Zhong [3] proposed a robust method as follows: Firstly, select special blobs according to the features of people images after background estimation. Secondly, use segmented blobs as the input of a people counting classifier. Then train the blobs to predict the number of people in each input image. These methods [2, 3] are time consuming due to the complex computation. Besides, to avoid segmenting the crowd people image, the number of people is estimated according to some allowable solutions [6, 7]. All of them are imprecise.

In this study, we have developed a people counting system based on head detection using a novel descriptor (BFR). The BFR descriptor segments a normal person into three parts: invariable area including the torso, variable area including the leg, the third part is the head area. Besides, we count people's number using MEA (Much Evidence Algorithm) based on color analysis in image sequences. Results show that our method is accurate and is free from the influence of people wearing various hats or walking with each other.

III. THE PROPOSED PEOPLE-COUNTING SYSTEM

A. System Configuration

While Fig 1 shows the configuration of our system, a color video camera is set on the ceiling of the gate with

θ degree towards the ground so that people will be observed in the ROI (Region of interest). We set two counting-lines (In and Out) according to various environments which can help us to make more accurate decisions about the directions of the path. When people pass through the In-line, we detect and track the head of them, and after they passing through the Out-line, the number of people is updated. For the out number, do the same.

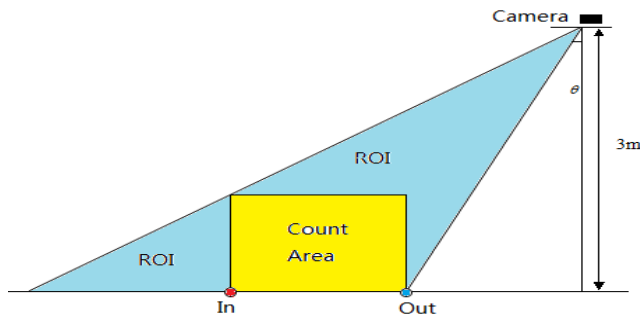


Fig 1. System Configuration (Where θ equals 30 degrees)

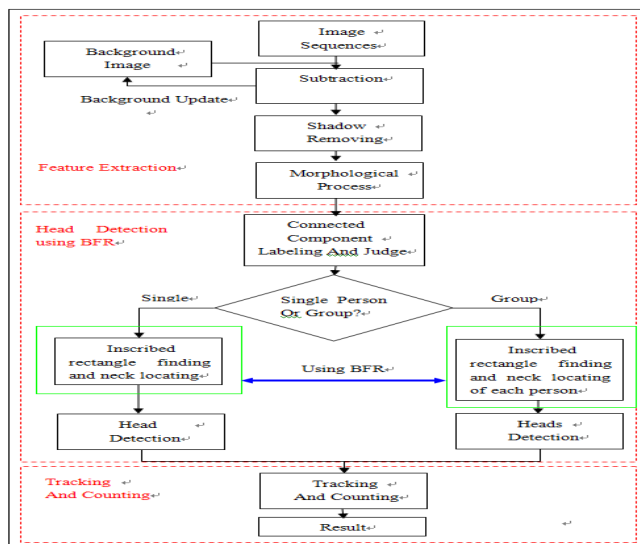


Fig2. Flow diagram of algorithm

Fig 2 shows a flow diagram of the proposed method. The whole system is composed of three parts: feature extraction (for moving people), head detection using BFR, tracking and counting. For the first part, we get the binary image of moving people exactly. For the second part, we use the key algorithm of head detection to identity head location whatever one or more people. For the last part, we track the head area and count the number.

B. Feature Extraction

There has been an amount of research in feature extraction. Several background update models were proposed such as region background model, Gaussian model, non-parametric model, codebook model and so on. Since the region background model [8] is very fast and satisfactory for our application, we choose it as our background update method.

Furthermore, we develop the method mentioned in [9] to remove shadows cast of people after background subtraction in the ROI. Because shadow often makes significant change in intensity with little change in chromaticity, the normalized chromaticity based RGB color space is used to detect shadow

of human objects. The formula as follows:

$$C_r = \frac{R}{R + G + B} \quad (1)$$

$$C_g = \frac{G}{R + G + B} \quad (2)$$

$$C_b = \frac{B}{R + G + B} \quad (3)$$

We define C_r, C_g, C_b as the normalized chromaticity value for the (i, j) th pixel of background image, and C_r, C_g, C_b as the normalized chromaticity value for the (i, j) th pixel of foreground image. $G_b(i, j)$ is the gray value for the (i, j) th pixel of background image, $G_f(i, j)$ is the gray value for the (i, j) th pixel of foreground image. The formula as follows:

$$C = \text{Max}\{|C_r - C_r|, |C_g - C_g|, |C_b - C_b|\} \quad (4)$$

$$G(i, j) = G_b(i, j) - G_f(i, j) \quad (5)$$

$$P(i, j) = \begin{cases} \text{Shadow,} & \text{if } (C < Th2) \text{ and } (G > Th1) \\ \text{Object,} & \text{if } (C > Th2) \text{ and } (G > Th1) \\ \text{Background,} & \text{else} \end{cases} \quad (6)$$

In order to remove noise and fill in small holes of the extracted human image, a fast modified masking method proposed in [6] is used instead of morphological processing. The whole processing of feature extraction is shown as Fig 3.

C. Multiple-Human Segmentation Based on Head Detection Using BFR

Multiple-human image segmentation is to divide touching persons into individuals. Because the head is visible and stable most of the time while we are walking, it is easy to be tracked and counted in the people flow-counting system. If we find the head position, we will solve the merge-split case that people touching with each other.

Zui Zhang proposed a method [10] using the XYZ and HSV color spaces to analyze the color of hair and skin by sets of Gaussian mixtures models. Because hair and skin are various, this method cannot get head image accurately.

In the proposed scheme, we will introduce head detection method based on BFR descriptor in details to overcome two difficult problems. One is to separate people image exactly, and the other is to avoid the influence of people wearing various hats. The whole process includes two parts as follows:

Part1. Vertical Projection Algorithm (VPA) for the segmentation of touching people images.

In our counting system, the walking people are almost facing to the camera in the ROI. The common gesture is shown in Fig 4(a).

We can see three touching people in Fig 4(a). In order to divide them into individuals, a simple algorithm named VPA is proposed as follows:

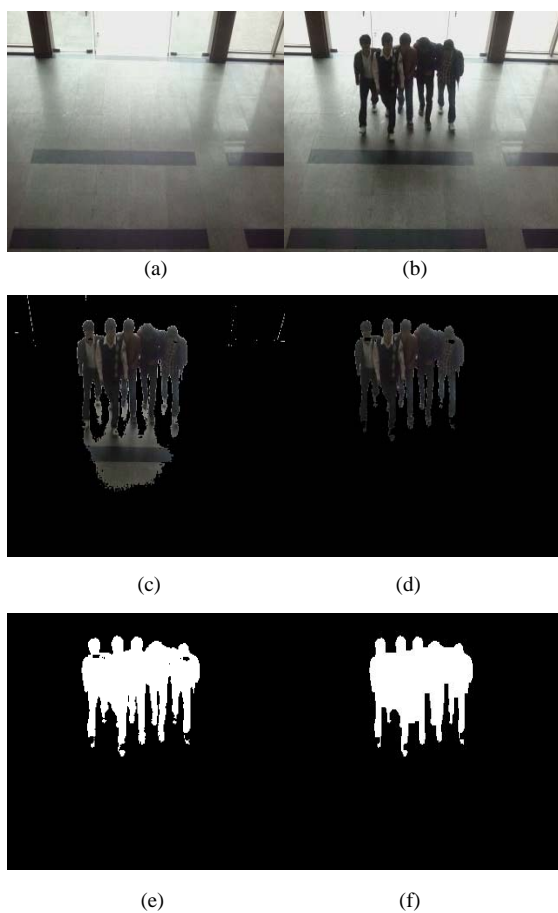


Fig3. Results of feature extraction

(a) background image, (b) the current frame, (c) result of background subtraction, (d) shadow remove, (e) binary image, (f) the extracted target after masking method

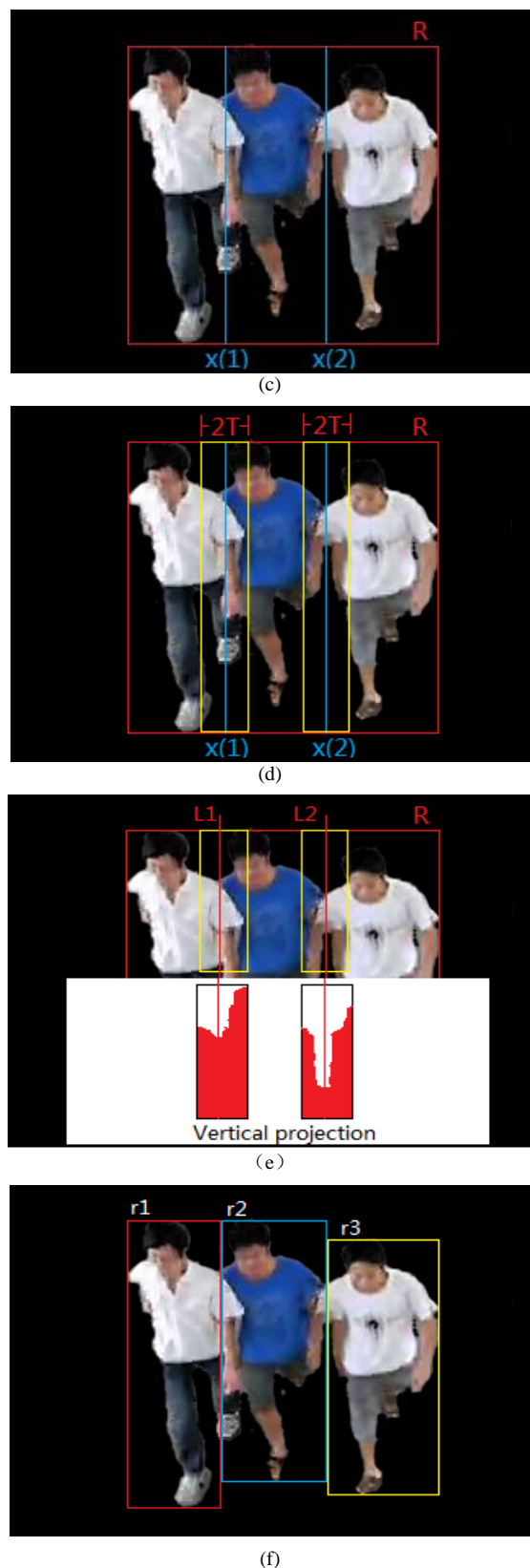


Fig4. A crowd people image after background subtraction

Step1. The enclosing rectangle of the people image is got after background subtraction, and named as R. The result is shown in Fig 4(b). We define the width of the enclosing rectangle for a single person image as r , that is to say the typical width of a single person image is r pixels per people. According to the value of r and the width of R , we can get the number of people and the lines to divide the crowd into

individuals using formula (7). The result is shown in Fig 4(c).

$$\begin{aligned} \text{Num} &= \text{WR}/r > (\text{int})(\text{WR}/r) ? (\text{int})(\text{WR}/r) + 1 : (\text{int})\text{WR}/r; \\ x(1) &= \text{WR}/\text{Num}; \\ x(2) &= 2 * \text{WR}/\text{Num}; \\ x(i) &= i * \text{WR}/\text{Num} \quad (1 \leq i \leq 6) \end{aligned} \quad (7)$$

$x(i)$ means the i th segmentation line, WR means the width of R .

But at this stage, the segmentation line is imprecise so we call it estimated line.

Step2. Get accurate segment lines.

In order to get accurate segment lines but not the estimated results, a simple method based on vertical projection is proposed. A threshold T is set to get some special detection areas. These areas include the accurate position which can divide the crowd image into individuals accurately. They are shown in Fig 4(d). The yellow rectangles are special detection areas. In order to avoid the influence of legs, we get the half of the yellow rectangle as the needed which dose not include legs.

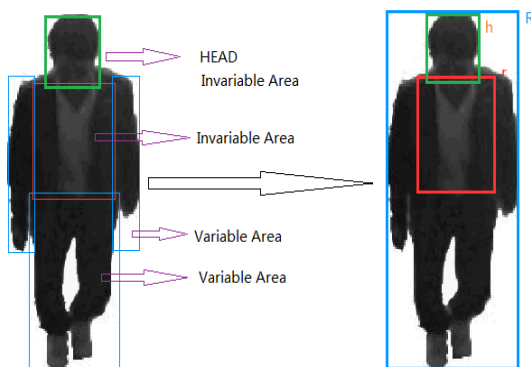


Fig5. Body feature rectangles

After getting the detection area, we get the vertical projection image and then get the accurate segment lines from it. The segment lines are drawn in red in Fig 4(e). Each line is the minimum value of the vertical projection values. Then, we get the individuals which are shown as different rectangles in Fig 4(f).

Part2. Head detection based on Body Feature Rectangles (BFR).

For a moving person, his legs and arms are variable, but his head and torso are stable. So we divide the human body into two areas in Fig 5. Head and torso belong to the invariable area while legs and arms belong to the variable area. In order to describe the body feature what we need, we make the enclosing rectangle of a walking people as R in Figure 5. R includes the invariable area and the variable area. If we remove the variable area including legs and arms, we will get the invariable area including the torso and the head. We use a special inscribed rectangle r in red color and a small rectangle h in green color to describe the torso and the head area respectively in Fig 5. We find that the boundary between the head and the torso is the neck position. As soon as the neck position is located, we can get the head position easily.

In order to get the head position, the whole processing is

shown in Fig 6.

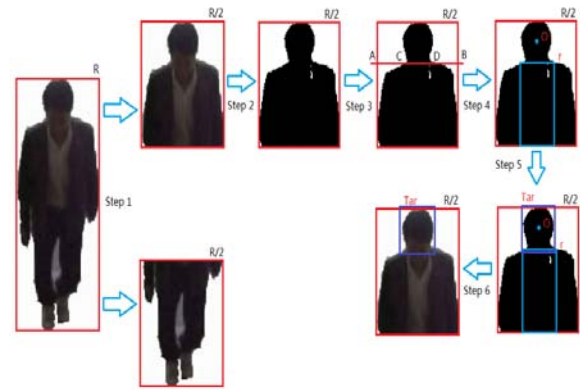


Fig6. The whole processing to get the head position

The original people image is divided into two parts in step1 in Fig 6. The top part includes the invariable area and the other includes the variable area. The binary image of the invariable area is got in step 2 in Fig 6.

From the body geometry, we know that the width of our neck is smaller than $R/2$. There is a great difference between the width of the torso area and the neck area. Based on this theory, we can get the key line AB and the joint, C and D , in step 3 in Fig 6. Then the inscribed rectangle r is got easily in step 4. Besides, the invariable area is divided into two targets

by the key line AB . The target in the top part is our head. The centroid of the head target is counted by the formula (8)-(10).

$$f(x, y) = \begin{cases} 1 & \text{target} \\ 0 & \text{background} \end{cases} \quad (8)$$

$$x_0 = \frac{\sum x f(x, y)}{\sum f(x, y)} \quad (9)$$

$$y_0 = \frac{\sum y f(x, y)}{\sum f(x, y)} \quad (10)$$

Last, the enclosing rectangle of the head is got in step 5 in Fig 6. The blue rectangle named Tar is obtained in step 6.

We test the image resulted in Part 1, and the test result is shown in Fig 7. We can see that three heads are located accurately using different rectangles.



Fig7. The result of head detection

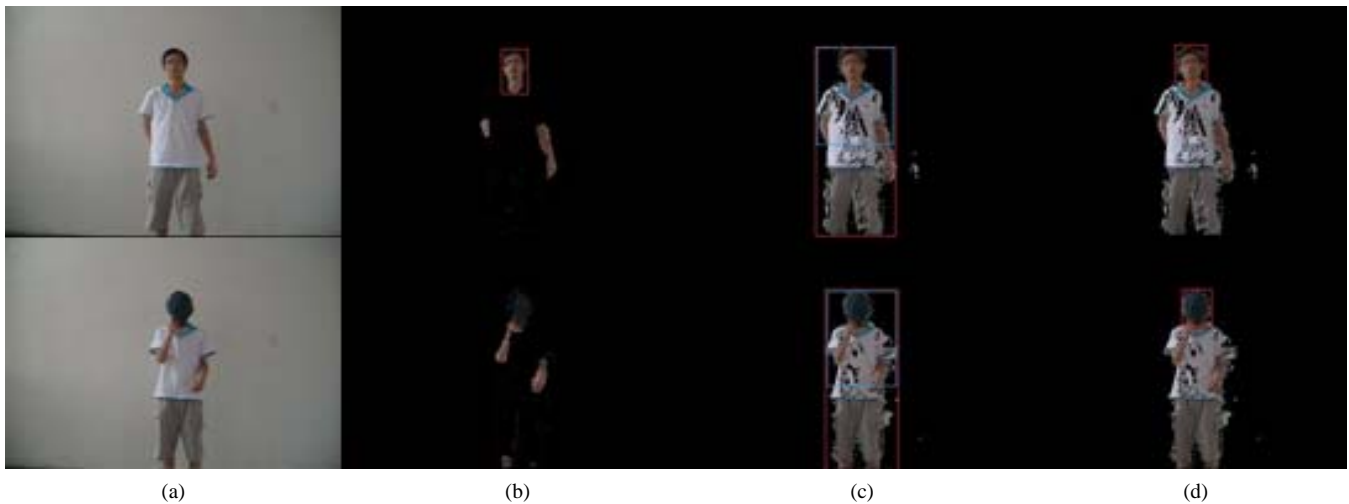


Fig 8. The results of proposed method and the method [10] (a)the original images(b)the result of the method [10](c-d)the result of proposed method

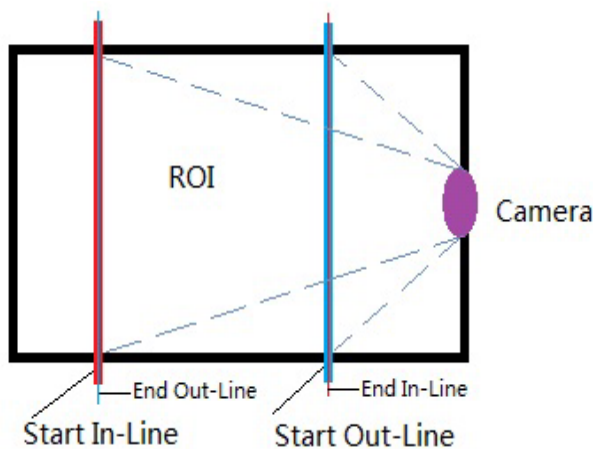


Fig9. Counting system image

Because of using the body's geometry features instead of color information, we can get the head position precisely when people are wearing hats. Compared with the method [10], the result is shown in Fig 8. The result of method [10] is shown in Fig 8(b) and our result is shown in Fig 8(c-d). We can see both of the two methods are effective when people don't have a hat. However, the method [10] is invalid while the proposed method is still effective when the face of person is covered by a hat.

D. Tracking and Counting Using MEA

In our method, we use a novel mean shift based tracking method [11] to track the head of person. This tracking method utilized the probability density distribution of the target gradient angle as the feature and constructed a similarity function that can be optimized by mean shift method. The result shows that it is very fast and satisfactory for our application.

For the counting algorithm, we use two count-lines judgments. It is shown in Fig 9. We detect and track the head when someone passes through the Start In-Line. If he passes through the End In-Line, we update the number of people who pass through the gate. For the out-people, we do the same.

Besides, if we just detect one frame to get the final number of people, it will be imprecise. So we use a new algorithm called MEA (Much Evidence Algorithm) to get the final number.

In MEA, we test 7 continuous frames instead one frame. We detect them separately and get heads' number as N_i in i th frame. We compare the value of N_i and set a threshold T_0 to get the final number of people. If there are at least T_0 frames which have the same value n , the final number of people equals n . We count the real number as N_f . The Pseudo code of MEA as follows:

define the number of head in each frame is N_i

$0 \leq i \leq 7$

compare the value of N_i

count the number of the same value in N_i as N_c

if $N_c \geq 5$

$N_f = N_c$

The robustness of MEA and the normal method are tested in the case of 100 frames. The result is shown in Fig 10.

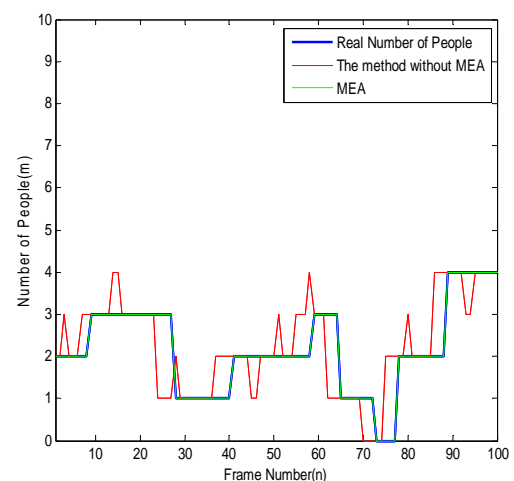


Fig 10. The robustness of MEA and the normal method

From Fig 10, we can see the blue curve and the green curve are closely while the red one has many peaks and valleys of wave. The reason is that the normal method counts the number of people in one frame. If some people are lost in one frame, the final number of people will be false. To the contrary, the MEA counts the final number using 7 frames instead of one frame, so it is precise and robust.



Fig11. The Result of MEA



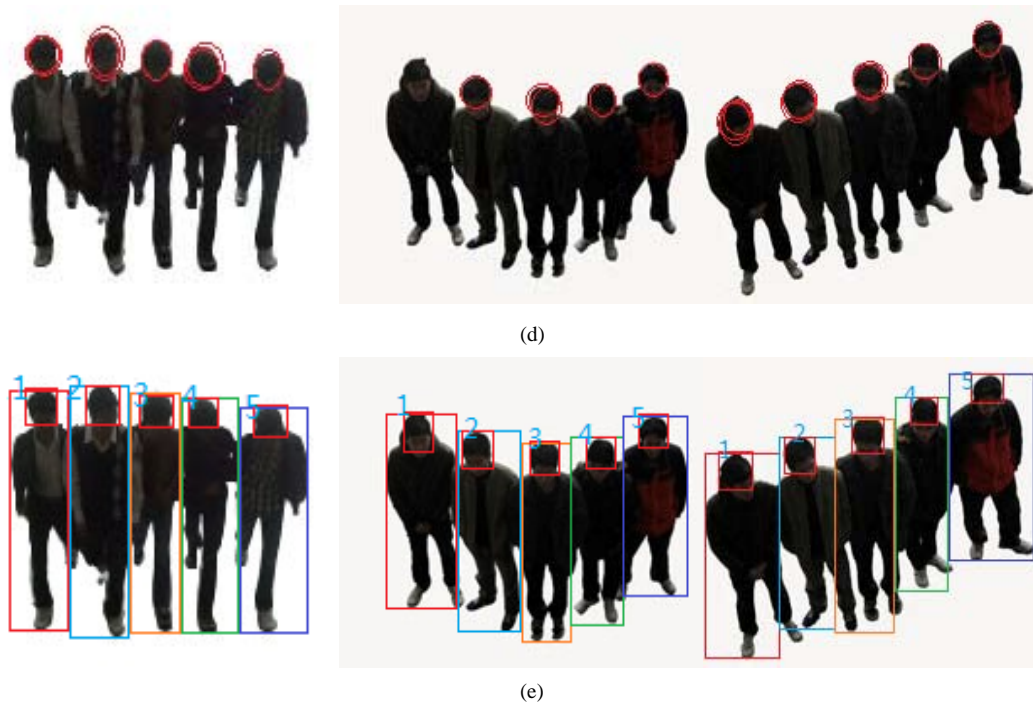


Fig 12. Results of different methods

(a) original images (b) results of method [10] (c) results of method [2] (d) results of method [12] (e) results of the proposed method

We choose 7 frames in Fig 11. We can see 6 frames are counted correctly, and the number of heads in each frame is 3, hence the final number of people is 3.

IV. EXPERIMENTS AND RESULTS

We test our system on PC (Intel(R) Celeron CPU 1.80GHz, 2GB memory) and Windows 7 operation system. The video size is CIF (320*240).

We test several images in different cases using different methods in Fig 12.

In Fig 12, the method [10] can not get all the heads in Fig 12(b) because our skin color is closely to our clothes and hairs in some bad light conditions. The method [2] also lost several targets for its learning system is not able to contain all the head-shoulder models. The method [12] gets head areas using Hough translation algorithm. It can not get all head positions since our heads may be not circular for our hair or hats. To the contrary, the proposed method is effective in Fig 12(e).

We compared several methods to get the heads number in four cases in 1000 frames. The result is shown in Fig 13. In order to analyze the robust of the proposed method using MEA, we compared four methods in 500 frames. The result is shown in Fig 14. Besides, we choose some frames to present the result in Fig 15.

In the Fig 13, case one means that people are walking separately without overlapping, case two means that people are wearing hats separately without overlapping, case three means that people are walking together without wearing hats, case four means that people are walking together and wearing hats. We can see that the method [2] is better than the method [10] and method [12]. The accurate ratio is about 91.9%. The method [12] can not get an ideal result when people are walking together and wearing hats for our head is not a circle in that case. The method [10] is effective when people are walking separately and facing to the camera as well. If our

heads are covered with hats or other things, it will not get our heads information. To the contrary, the proposed method plays well in different cases whatever people are walking together or wearing hats and the average accurate ratio is about 95.75%.

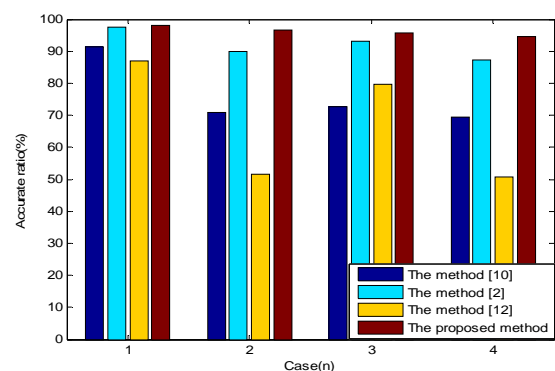


Fig13. Results of different methods

In the Fig 14(a), the red line is the real number of people and the blue one is the counting result of the method [2]. The blue one has several peaks and troughs which are different from the real number. This method gets the wrong number at the peak or trough because it counts the number just in one frame. So are the reason of Fig 14(b) and Fig 14(c). Except that, because of different color of skin and the influence of our hats, the blue line has so many peaks and troughs in Fig 14(b). That's to say the method [10] maybe make many errors in counting numbers. The method [12] which based on the Hough transform theory is the worst one among them. Our heads can not like circles when we have hats or have various hairs so that the blue line has many peaks and troughs.

Compared with them, the proposed method has a few peaks. Because the real number is resulted by 7 frames continuously, the proposed method will not get a wrong number at each peak in blue line. We can see that the blue line and the red one are the same except a little offset.

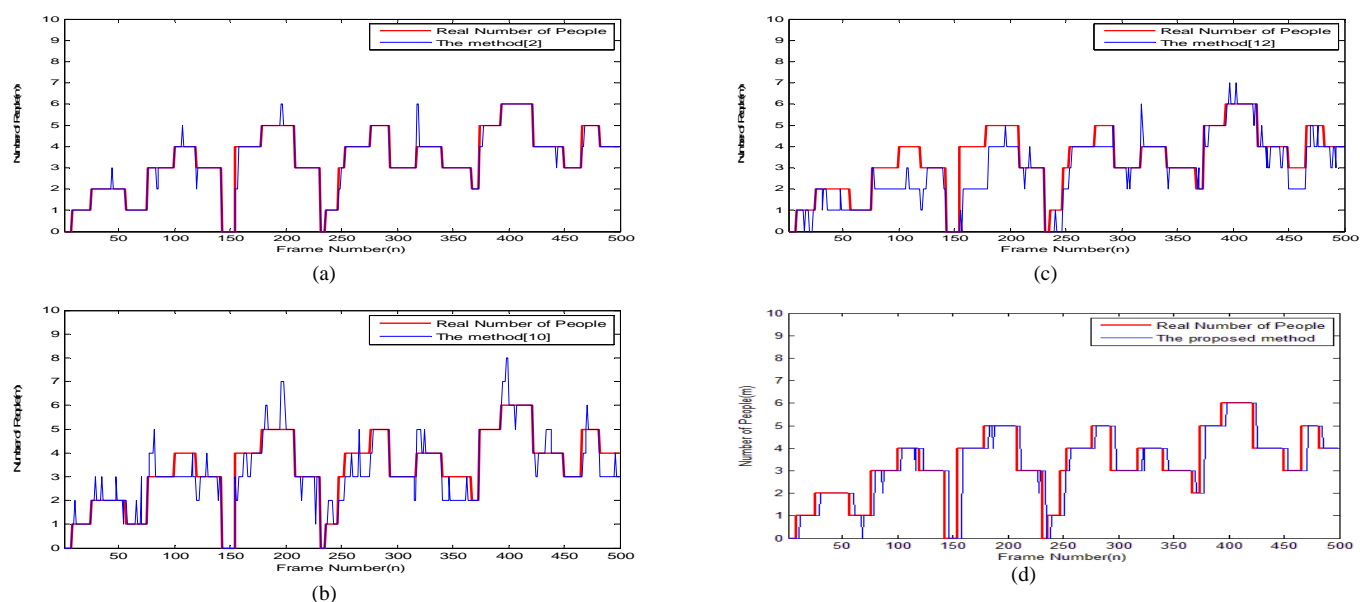
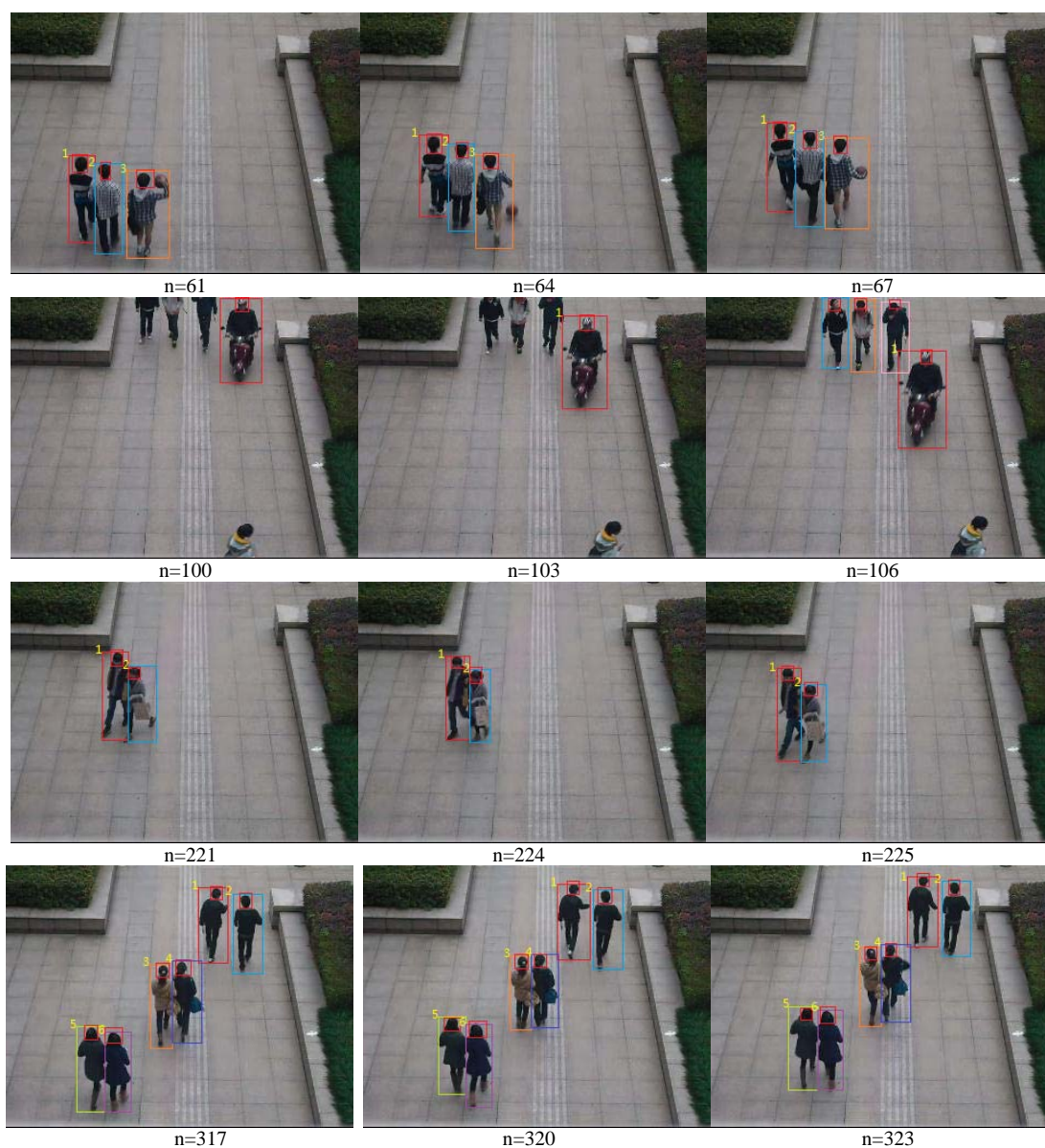


Fig.14 The robust of different methods (a)the method[2](b)the method[10](c)the method[12](d)the proposed method



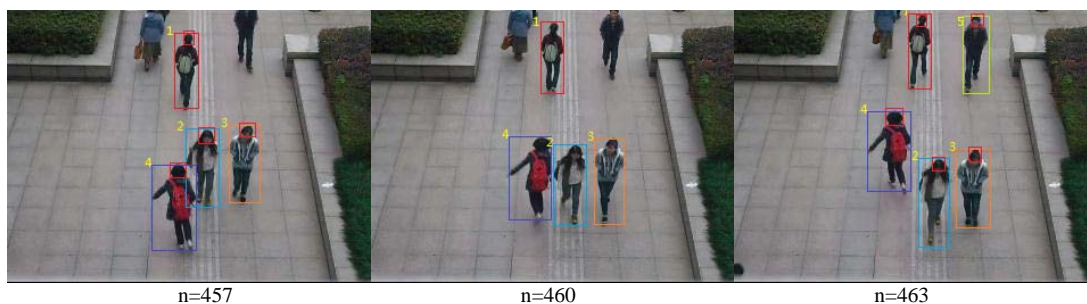


Fig.15 Results of counting

In Fig 15, we can see the proposed method is effective in different cases whatever people touch with each other or not. Moreover, a motorcyclist is detecting accurately in some frames.

The experimental result of head detection for multiple-people images is shown as table 1. The counting result is shown in table 2. We test in a normal environment: the light intensity is invariable during the whole time. We test time cost of two methods, one is the proposed method, and the other is the method [10]. The method [12] is time consuming

so that we cannot consider it. Our method is shown as Fig16 (a)

The method [10] is shown as Fig 16(b). Compared with Fig 16(b), when the number of overlapping-people is 6, both of two methods spend time closely. But when the number is less than 6, we see that the method [10] is time consuming and the proposed method is effective. According to the analysis of results, if the number of crowd is less than 6, our counting-system has stronger robustness and accuracy in different time and scene.

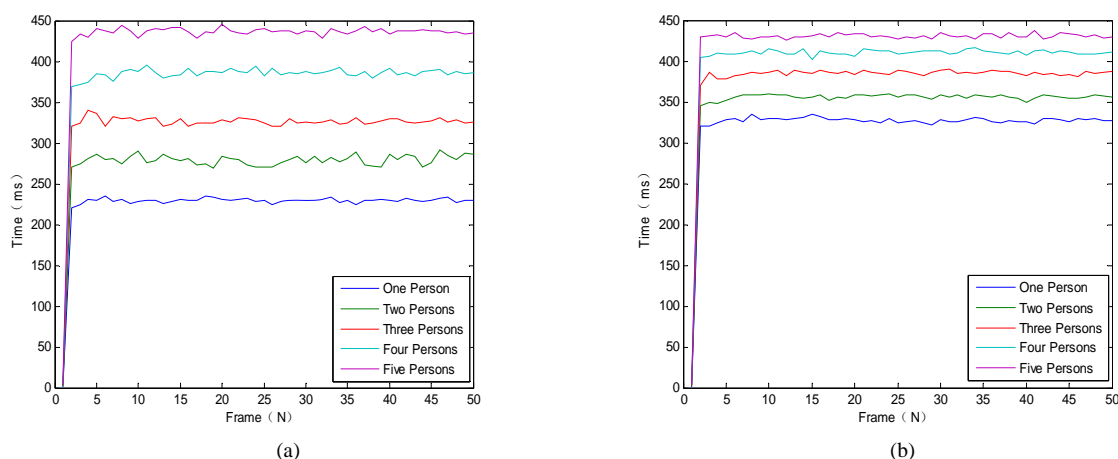


Fig16. Results of method [10] and the proposed method (a) the proposed method; (b) method [10]

TABLE1. EXPERIMENTAL RESULTS OF HEAD DETECTION FOR MULTIPLE-PEOPLE IMAGES

NO. image	Multiple People	NO. Frames	Heads detection	Time cost(ms)	Average Accuracy rate
1		53	53	22.935	100%
2		49	98	27.910	100%
3		57	168	32.676	98.2%
4		60	232	38.537	96.7%
5		55	263	43.653	95.6%

TABLE2. EXPERIMENTAL RESULTS OF PEOPLE COUNTING

Video	In Number	Out Number	Count In	Count Out	Average Accuracy rate
Scene 1	28	31	27	29	94.915%
Scene 2	52	46	48	45	94.898%
Scene 3	76	68	72	63	93.75%

V. CONCLUSIONS AND FUTURE WORK

Experimental results show that our algorithm of people-flow counting is fast and precise. Our system can run in real-time.

We do not solve the problem that the number of people is more than six. Moreover, if people crowd enough, our method will be imprecise. We will consider completely in future.

REFERENCES

- [1] Yu Ming, Liu Wei, A new method for passenger flow counting system based on surveillance video, in: Proceedings of the 2nd International Conference on Intelligent Networks and Intelligent System, pp. 453-456, 2009.
- [2] Huazhong Xu, Pei Lv, Lei Meng, A people counting system based on head-shoulder detection and tracking in surveillance video, in: 2010 International Conference on Computer Design and Applications, pp.V1394-V1398, 2010.
- [3] Weizhong Ye, Yangsheng Xu, Zhi Zhong, Robust people counting in crowded environment, in: 2007 IEEE International Conference on Robotics and Biomimetics, pp.1133-1137, 2007.
- [4] Tsong-Yi Chen, Zhi-Xian Chen, An intelligent people-flow counting method for passing through a gate, in: 2006 IEEE Conference on Robotics, Automation and Mechatronics, pp.1969-2011, 2006.
- [5] P.H.Tu, Rittscher J, Perera A, Krahnstoever N, Detecting and counting people in surveillance applications, in: IEEE International Conference on Advanced Video and Signal Based Surveillance Proceedings of AVSS 2005, pp.306-311, 2005.
- [6] Hyun Hee Park, Hyung Gu Lee, Seung-In Noh, Jaihie Kim, An area-based decision rule for people-counting systems, in: Lecture Notes in Computer Science, pp.450-457, 2006.
- [7] Enwei Zhang, Feng Chen, A fast and robust people counting method in video surveillance, in: Proceedings of 2007 International Conference on Computational Intelligence and Security, pp.339-343, 2007.
- [8] Tao Lei, Yandyu Fan, Liandong Li, The algorithm of moving human body detection based on region background modeling, in: Proceedings of 1st International Symposium on Computer Network and Multimedia Technology, 2009.
- [9] Stephen J.McKenna, Sumer Jabri, Zoran Duric, Azriel Rosenfeld, Harry Wechsler, Tracking groups of people, in: Computer Vision and Image Understanding, pp.42-56, 2000.
- [10] Zui Zhang, Hatice Gunes, Massimo Piccardi, An accurate algorithm for head detection based on XYZ and HSV hair and skin color models, in: 2008 IEEE International Conference on Image Processing, pp.1644-1647, 2008.
- [11] Wang Changjun, Zhang Li, Mean shift based orientation and location tracking of targets, in: Proceedings of 2010 6th International Conference on Natural Computation, pp.3593-3596, 2010.
- [12] Hong Liu, Yueliang Qian, Shouxun Lin, Detecting persons using Hough circle transform in surveillance video, in: 2010 International Conference on Computer Vision Theory and Applications, pp.267-270, 2010.